Setting the basis of best practices and standards for curation and annotation of logical models in biology

OXFORD UNIVERSITY PRESS | Briefings in Bioinformatics

Setting the basis of best practices and standards for curation and annotation of logical models in biology – Highlights of the [BC]2 2019 CoLoMoTo/SysMod Workshop

Journal:	*Briefings in Bioinformatics*
Manuscript ID	BIB-19-0501.R1
Manuscript Type:	Method Review
Date Submitted by the Author:	20-Feb-2020
Complete List of Authors:	Niarakis, Anna; Universite d'Evry-Val-d'Essonne, Department of Biology, GenHotel, Genopole Kuiper, Martin ; Norwegian University of Science and Technology, Department of Biology Ostaszewski, Marek; Universite du Luxembourg, Luxembourg Centre for Systems Biomedicine Sheriff, Rahuman S; European Bioinformatics Institute, BioModels Casals-Casas, Cristina; SIB Geneva, Swiss-Prot group, SIB Swiss Institute of Bioinformatics, 1 Michel Servet, 1211 Thomas, Paul; University of Southern California, GO, (GO-CAM/Noctua) Thieffry, Denis; IBENS, Computational Systems Biology Team, CNRS, INSERM, École Normale Supérieure, PSL Research University; National University of Singapore Life Sciences Institute Freeman, Tom; Roslin Institute, Genetics and Genomics Touré, Vasundra; NTNU, Department of Biology Noël, Vincent; Institut Curie, U900 INSERM / Mines ParisTech Stoll, Gautier; Centre de Recherche des Cordeliers, INSERM UMRS1138, Sorbonne Université, Université Paris Descartes, Université Paris Diderot; Gustave Roussy, Metabolomics and Cell Biology Platforms Rodriguez, Julio; Heidelberg University, Faculty of Medicine, & Heidelberg University Hospital; RWTH Aachen University, JRC for Computational Biomedicine, Faculty of Medicine Naldi, Aurelien; IBENS, Computational Systems Biology Team, CNRS, INSERM, École Normale Supérieure, PSL Research University Oshurko, Eugenia; ENS de Lyon, UCBL, CNRS, LIP, Xenarios, Ioannis; UNIL, CHUV, Ludwig Institute of Cancer Research Soliman, Sylvain; Inria Centre de Recherche Saclay Ile-de-France, Équipe Lifeware Chaouiya, Claudine; Aix-Marseille Universite, CNRS, Centrale Marseille, I2M Helikar, Tomas ; University of Nebraska-Lincoln, Department of Biochemistry Calzone, Laurence; Institut Curie, Paris, France, Bioinformatics
Keywords:	model reusability, annotation standards, logical modelling, biocuration, reproducibility

Setting the basis of best practices and standards for curation and annotation of logical models in biology – Highlights of the [BC]2 2019 CoLoMoTo/SysMod Workshop

Anna Niarakis, Martin Kuiper, Marek Ostaszewski, Rahuman S Malik Sheriff, Cristina Casals-Casas, Denis Thieffry, Tom C. Freeman, Paul Thomas, Vasundra Touré, Vincent Noel, Gautier Stoll, Julio Saez-Rodriguez, Aurelien Naldi, Eugenia Oshurko, Ioannis Xenarios, Sylvain Soliman, Claudine Chaouiya, *Tomáš Helikar and *Laurence Calzone

*Indicates equal contribution

Authors Affiliations

Anna Niarakis GenHotel, Univ. Évry, Université Paris-Saclay, Genopole, 91025 Évry, France
Martin Kuiper, Department of Biology, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
Marek Ostaszewski- Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 6 Avenue du Swing, L-4367 Belvaux, Luxembourg
Rahuman S. Malik Sheriff, European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom
Cristina Casals-Casas, Swiss-Prot group, SIB Swiss Institute of Bioinformatics, 1 Michel Servet, 1211 Geneva 4, Switzerland
Paul Thomas – GO, (GO-CAM/Noctua), USC, USA
Denis Thieffry, Computational Systems Biology Team, Institut de Biologie de l'École Normale Supérieure (IBENS), CNRS, INSERM, École Normale Supérieure, Université PSL, Paris, France
Tom C. Freeman, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Edinburgh, EH25 9RG, United Kingdom
Vasundra Touré, Department of Biology, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
Vincent Noël, Institut Curie / U900 INSERM / Mines ParisTech, France
Gautier Stoll, Team of Metabolism, Cancer & Immunity, Centre de Recherche des Cordeliers, INSERM UMRS1138, Sorbonne Université, Université Paris Descartes, Université Paris Diderot, Paris, France; Metabolomics and Cell Biology Platforms, Gustave Roussy Cancer Campus, Villejuif, France

Julio Saez-Rodriguez: Institute for Computational Biomedicine, Heidelberg University, Faculty of Medicine, & Heidelberg University Hospital, Bioquant - Room 464 Im Neuenheimer Feld 267, 69120 Heidelberg, Germany
JRC for Computational Biomedicine, RWTH Aachen, Faculty of Medicine, Aachen, Germany
Aurélien Naldi, Computational Systems Biology Team, Institut de Biologie de l'École Normale Supérieure (IBENS), CNRS, INSERM, École Normale Supérieure, Université PSL, Paris, France
Eugenia Oshurko, Univ Lyon, EnsL, UCBL, CNRS, LIP, F-69342 LYON Cedex 07, France
Ioannis Xenarios, CHUV Agora Center, UNIL, Ludwig Institute of Cancer Research, Rue du Bugnon 25A, Lausanne, Switzerland
Sylvain Soliman, Lifeware group, Inria Saclay-Île de France, 91120 Palaiseau, France
Claudine Chaouiya, Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France
Tomas Helikar, Department of Biochemistry, University of Nebraska-Lincoln Lincoln, NE, USA
Laurence Calzone, Institut Curie, PSL Research University, Mines ParisTech, Inserm, U900, F-75005, Paris, France

Abstract

The fast accumulation of biological data calls for their integration, analysis and exploitation through more systematic approaches. The generation of novel, relevant hypotheses from this enormous quantity of data remains challenging. Logical models have long been used to answer a variety of questions regarding the dynamical behaviours of regulatory networks. As the number of published logical models increases, there is a pressing need for systematic model annotation, referencing and curation in community-supported and standardised formats. This article summarizes the key topics and future directions of a meeting entitled "Annotation and curation of computational models in biology", organized as part of 2019 [BC]2 conference. The purpose of the meeting was to develop and drive forward a plan towards the standardised annotation of logical models, review and connect various ongoing projects of experts from different communities involved in modelling and annotation of molecular biological entities, interactions, pathways and models. This article defines a roadmap towards annotation and

curation of logical models, including milestones for best practices and minimum standard requirements.

Keywords: biocuration, logical modelling, reproducibility, model reusability, annotation standards

Introduction

Reproducibility of research findings constitutes a key concern of the scientific community as multiple reports show that published results in various scientific domains cannot be replicated [1]. In the field of computational systems biology, where scientists combine prior knowledge based on experimental evidence and computational approaches, the reproducibility of results can be fostered through the use of consensual practices and standards, extensive annotation, code sharing, as well as by depositing of the resulting models in dedicated repositories. Logical (or logic) models (Boolean, multivalued, or other variants) have been widely used for studying and analysing in-depth regulatory mechanisms and biological processes for which kinetic data are scarce. Some repositories for this type of models exist already, including GINsim repository [2] and Cell Collective, a platform for building, analysing and visualising models [3,4].

In the GINsim repository, one can find models built with the software GINsim and used for simulations in peer-reviewed articles. Models are stored in their zginml format and a summary along with a link to the supporting scientific article are provided. In Cell Collective , models have been manually curated by re-construction, re-simulation and analysis to ensure that their dynamics correspond to published results. Efforts are further made to include logical models in BioModels, a repository of mathematical models of biological and biomedical systems [5]. Annotation practices, accuracy and reproducibility checks made by the BioModels team will facilitate consistent quality control of these models.

To facilitate exchanges of logical models and communication between tools, previous work by the CoLoMoTo consortium and Systems Biology Markup Language (SBML) teams focused on standardisation of model formats by developing a specific package of the Systems Biology Markup Language level 3 (SBML L3) [6], SBML-*qual* [7,8]. However, specific minimum requirements for the annotation and level of curation of logical models remain to be defined. Even when results are reproducible, models often fail to be reusable because of the lack of explicit referencing to the sources used for their construction (organism, experimental setting and type of data, published manuscript sources, identifiers to relevant database entries, etc.).

To address the pressing need to propose and develop best practices and standards in annotation and curation of logical models in biology, Anna Niarakis, Laurence Calzone and Tomáš Helikar (representatives of the CoLoMoTo [9] and SysMod [10] communities) organized a workshop in the context of the [BC]^2 conference recently held in Basel [11], with the aim to bring together logical modellers and curators. The workshop, entitled "Annotation and curation of computational models in biology" [12] is the most recent of a series of workshops organised by the logical modelling community over the past years, in the context of prominent international conferences such as ECCB 2014 (Strasbourg, France), ICSB 2015 (Singapore), ECMTB 2016 (Nottingham, UK), [BC]2 2017 (Basel, Switzerland), ECCB 2018 (Athens, Greece).

The meeting was divided into four sessions highlighting key challenges of the modelling community (Figure 1), starting with curation platforms and model repositories. In particular, the need for establishing annotation criteria, quality control checks and the use of a common repository were extensively discussed. The following session focused on recent methodological advancements to analyse logical models to ensure interoperability and reusability, and lastly, the afternoon sessions were focused on integrative approaches and tools.

In **Table 1**, we have summarized briefly the topics discussed in each session. The presentations were followed by an extensive discussion between all speakers and participants on three key topics:

Reproducibility, i.e., the ability to replicate scientific results using the same model.

Reusability, i.e., the ability to reuse an existing model using transparent biocuration processes, extensive annotations and references that increase the model's liability.

Interoperability, i.e., the ability to analyse the same model with multiple tools due to the use of standard formats.

Model curation and annotation, and available repositories

The first session was dedicated to annotation and curation approaches, together with relevant repositories, including the presentation of curation approaches and tools for the development of Boolean models for colon cancer and molecular causal interaction statements, the introduction to the complementary platforms BioKB [13] and MINERVA [14], followed by that of the BioModels repository. An example of an atherosclerotic plaque formation model demonstrated the necessity of proper annotation for optimal model-based predictions. The first session highlighted the necessity to annotate prior knowledge networks (PKNs) and logical models accurately for reusability, and enrich them with knowledge from heterogeneous resources to avoid potential ambiguities (e.g., UniProtKB [15], SIGNOR [16], HGNC [17], GO [18], REACTOME [19]).

Martin Kuiper (DrugLogics team, NTNU) presented work on a set of Colon Cancer logical models named CASCADE (CAncer Signaling CAusality DatabasE, [20]), and the development of a novel curation interface named Visual Syntax Method (VSM, [21]), which enables the curation of biological network information that includes causal molecular relationships. The VSM interface was tested extensively to annotate the full collection of experimentally analysed

DNA binding transcription factors for human, mouse and rat [22], and is now being implemented in *a curation platform for causal interaction statements* [23]. Causal interaction statements are basic representations of regulatory interactions between two biological entities that can be efficiently extracted from the literature, provided that proper annotation tools and curation guidelines are provided.

Marek Ostaszewski (Luxembourg Centre for Systems Biomedicine) presented *BioKB and MINERVA: a workflow for curation and fast prototyping of annotated knowledge repositories* [13,14]. To construct graphical models of molecular mechanisms, one needs to i) extract entities, interactions and relevant annotations from the literature, ii) build a consistent graphical representation, and iii) review and parameterise the model. BioKB [13,24] is a platform initially designed for exploring text mining data, which currently allows combining human-provided and machine-identified elements and their interactions into "facts" – human-curated relationships, annotated with sentences, literature and recognized identifiers. As BioKB is not a diagram editor, the biocurator can focus only on the accuracy of the extracted facts. This model visualization step, however, can be complemented with the MINERVA Platform, which allows API-driven [25] conversion of a layout-less model into an editable diagram (SBGN-ML, [26]) that can be further used by various systems biology editors (e.g., CellDesigner [27], Newt [28], etc.). This way also the final step of the model curation workflow can be realised - a curated diagram can be exported to the chosen systems biology format, refined and parameterised. Moreover, such API-based conversion makes it convenient to include in bigger bioinformatic workflows.

Following the effort towards transparency of the different steps leading to model construction and the reusability of these models, Rahuman S. Malik-Sheriff (European Bioinformatics Institute (EMBL-EBI)) discussed how *Curation and annotation of models in BioModels repository promote reproducibility and reusability*. Established in 2005, BioModels provides

a platform to support sharing, easy accessibility and reproducibility of mathematical models of biological processes [5,29]. Models submitted to BioModels are verified and curated by expert in-house curators. In 2011, an effort was made to extend the standard to logical formalism and SBML-qual was defined [7,8], allowing the inclusion of logical models in the database. Following Minimum Information Requested In the Annotation of Models (MIRIAM) guidelines, curated models are encoded in the standard SBML format and semantically enriched with controlled vocabularies [30]. Model entities are linked to several data resources (e.g., UniProt [15], Ensembl [31], the NCBI Taxonomy Database [32]), as well as ontologies, such as Gene Ontology [18], ChEBI [33], Mathematical Modelling Ontology [34], Systems Biology Ontology [35], and Brenda Tissue Ontology [36]. Such annotations allow the unambiguous identification of model components and processes. BioModels currently hosts over 900 curated models, becoming the world's largest repository of curated models. BioModels team will soon start to systematically curate logical models. To date, however, only seventeen logical models, three curated and fourteen non-curated are included in the BioModels' collection.

Cristina Casals-Casas from (Swiss-Prot) presented *SysVasc Prior Knowledge Network (PKN): An example of biocuration for dynamical modelling*. As a case study, Cristina Casals-Casas and collaborators have built a PKN to allow dynamical modelling of atherosclerotic plaque formation [37]. The expert curation strategy was centred on regulatory interactions between biological entities (gene products, chemical compounds and processes) interacting with each other in a complex manner, and exhibiting conditional dependencies between co-regulators. Biological entities were defined using strictly controlled vocabulary terms, retrieved from UniProtKB, HGNC, ChEBI, or GO, among others. The resulting PKN includes 729 components linked by 3,406 interactions of which 1,841 are complex regulations encoded with logical operators, while 1,565 are simply activatory or inhibitory interactions. For each

component, they demonstrated how the description of complex signalling functions and their integration are essential to correctly predict activation state in health and disease states. Their work highlighted the essential role of expert curation to correctly identify and encode complex regulatory interactions from experimental literature. Failure to encode these relationships correctly can alter significantly the behaviour of the model and the derived predictions. Dynamical models should be fine-tuned by contextualization to the specific biological system under study, and for this, proper annotation and expert curation are essential.

Community standards development and interoperability/reusability of existing models

The second session of the meeting was dedicated to interoperability and reusability of models and provided examples using three different model applications. Novel dynamical analysis methods and a framework for Gene Ontology annotations for supporting model building were also presented. All these approaches take advantage of existing databases to assist modellers and automatise error-prone and cumbersome tasks, currently still often performed manually, in order to optimise iterative modelling.

Denis Thieffry (Ecole Normale Supérieure, Paris) presented novel methods for the *Computational verification of large logical models,* with an *application to the prediction of T cell response to checkpoint inhibitors*. A first approach enables the formalisation and automatic verification of validation criteria for whole models or defined subparts, thereby greatly facilitating model development and correction. A second approach consists in computing the impact of specific environmental or genetic perturbations on model dynamics by propagating their impact on model logical rules. These methods were applied to the analysis of the impact of T lymphocyte checkpoint inhibitors and their use was integrated and illustrated in the

CoLoMoTo Interactive Notebook [38] (presented by Aurelien Naldi in the afternoon session) to foster transparent and reproducible analyses.

Tom Freeman (Roslin Institute) presented a *graphical and computational model of the renal mammalian circadian clock*. A comprehensive graphical model of the circadian pathway was constructed using the modified Edinburgh Pathway Notation scheme (mEPN) [39] and used to analyze the diurnal pattern of gene expression in the mouse kidney [40] using a stochastic Petri net-based approach [41]. The model encapsulates the interactions between 69 molecular species and contains 2,013 components and 2,100 interactions. All pathway components are labelled using standard nomenclature (HGNC gene id), and any modifications to those components are explicitly stated in their labels. Moreover, proteins, genes and biochemicals are hyperlinked to online resources, e.g., NCBI gene, ChEMBL and interactions between components (process nodes) are annotated with publications providing supporting evidence. In this respect, models can also serve as descriptive diagrams of known events that can be easily evaluated and reused.

Reinforcing this idea, Paul Thomas introduced *Gene Ontology Activity Modeling*. Gene Ontology (GO) annotations are the most comprehensive structured representation of gene function and are widely used in the interpretation of genome-wide experimental data. However, because an individual GO annotation associates a single gene product with a single GO term, it is only a partial description of the gene function, which limits the expressiveness of annotations and their application in computational analysis of experimental data. To address this limitation, Thomas *et al.* have developed a novel framework, GO Causal Activity Modeling (GO-CAM), for linking multiple GO annotations into an integrated model of a biological system. GO-CAM supports modelling at multiple levels, from individual gene products to complex regulatory and metabolic pathways, and can be applied in network

analysis and systems biology modelling, or converted into standard GO annotations for traditional GO-based analyses. Paul Thomas further presented the Noctua Modeling Tool used by GO Consortium curators to create GO-CAM models, from existing GO annotations or from scratch.

Finally, Anna Niarakis (UEVE, University of Paris-Saclay) closed the session by introducing the *automated inference of annotated Boolean models from molecular interaction maps using CaSQ*. She proposed a methodology to convert complex molecular maps into computable logical models. Molecular interaction maps have emerged as a useful way of representing biological mechanisms, based on information mining and human curation [40]. Nevertheless, their static nature does not allow for *in silico* simulations. With Sylvain Soliman (INRIA, Paris-Saclay), they have developed CaSQ [42], a tool that infers preliminary Boolean rules based on the topology and semantics of the molecular interaction maps, transforming these maps to executable Boolean models. They used a state-of-the-art molecular interaction map for Rheumatoid Arthritis [43,44] as a case study, but the tool can handle various maps differing in size and complexity and supports the SBGN standard. CaSQ inferred models are encoded in SBML *qual*, while references, annotations and layout are retained, thereby facilitating interoperability and model reusability.

Tools and modelling platforms for dynamical analysis of logical models

The afternoon sessions highlighted the efforts of the community to develop methodologies and software that address issues of interoperability, reproducibility and reusability of modelling efforts. The level of annotation and the amount of curation are highly dependent on the modeller and on the capabilities of the existing tools to support this type of information in both human and machine-readable formats.

Tomáš Helikar (University of Nebraska-Lincoln) introduced *Cell Collective - Enabling accessible and collaborative construction and analysis of comprehensive and annotated models*. Cell Collective is a computational modelling platform for the collaborative construction, simulation, and analyses of large-scale dynamic (logical) models of biological and biochemical processes [3,4,45]. It contains nearly 100 public, peer-reviewed logical models of various biological and biochemical processes. To ease the reuse and expansion of existing models, every component and interaction is annotated to track the biological data used to build the model. Models in Cell Collective can be created either *de novo* or imported using SBML-*qual*. Models are accessible in Cell Collective, where they can be simulated and further developed, or can be downloaded in SBML *qual* format, including via its public API [46].

Gaultier Stoll (Centre de Recherche des Cordeliers, INSERM) and Vincent Noël (Institut Curie) presented *MaBoSS (Markovian Boolean Stochastic Simulator) ecosystem*. MaBoSS is a tool for simulating logical models with continuous-time Markov processes [47]. Stochastic simulations allow the computation of the probabilities of each state of the model over time. Over the years, MaBoSS was extended [48] and various tools were developed, including UPMaBoSS, enabling the study the dynamical behaviour of cell populations (including its size), and PhysiBoSS, based on an agent-based formalism where each agent is a logical model run with MaBoSS. A model of cell fate decision was used to showcase different ways of running the tools: through the command line, through the CoLoMoTo Jupyter interactive notebook, showing the interoperability of the tool, and using the python library *pymaboss* [49].

Vasundra Touré (DrugLogics group, NTNU) presented *The Minimum Information about a Molecular Interaction Causal Statement (MI2CAST): a set of guidelines for the annotation of molecular causal interactions* [23]. The NTNU group proposes MI2CAST as a standard for representing causal statements that can serve as a checklist that can be followed in curation processes for capturing the essential contextual information about a causal relationship, to

ensure clarity, uniformity and reusability of the data across resources. MI2CAST has been developed in collaboration with the International Molecular Exchange (IMEx) consortium [50] and Human Proteome Organization - Proteomics Standards Initiative (HUPO-PSI) [51]. The NTNU group has also implemented the MI2CAST guidelines and annotation terms in a prototype curation tool based on the VSM foundation [21], named causalBuilder [52].

Julio Saez-Rodriguez (Heidelberg University) focused on *Integrating knowledge and experimental data to build context-specific logic models*. The general pipeline involves obtaining existing prior knowledge on pathways from available public resources using OmniPath [52], building a logic model from this prior knowledge, and training it to data with tools such as CellNOpt (for targeted readouts [53]), PHONEMeS (for untargeted mass spectrometry [54]), and CARNIVAL (for gene expression, [54]). Regarding annotations, OmniPath provides information about localisation, function, disease relationships, proteins and complexes based on 36 resources. Collectively, Omnipath provides 2,200,000 annotation entries for 20,000 human proteins and 16,500 complexes and is available via a Python module, an R package, as a web service, or from Cytoscape [55,56].

Aurelien Naldi (Ecole Normale Superieure, Paris) presented *The CoLoMoTo Interactive Notebook,* which provides a unified environment to edit, execute, share, and reproduce analyses of Boolean and multi-valued models of biological networks. This framework combines the power of different software tools to ensure reproducibility and to reduce their learning curve. The CoLoMoTo Interactive Notebook currently eases access to GINsim, BioLQM [57], Pint [58], MaBoSS, and Cell Collective. More tools will be included in the future. Computational workflows can be edited through a web interface based on the Jupyter notebook, enabling the inclusion of textual annotations, along with the explicit code to execute, as well as the visualisation of the results. The framework is distributed as a Docker image with the tools ready

to use without any installation step besides Docker, which can run on Linux, macOS, and Microsoft Windows systems.

Lastly, Eugenia Oshurko (Ecole Normale Supérieure, Lyon) presented *KAMIStudio: an environment for biocuration of cellular signalling knowledge* [59] suitable for rule-based modelling languages, such as Kappa [60] and BioNetGet [61]. KAMIStudio environment is based on the KAMI biocuration framework that aims to decouple knowledge curation from model building [62]. The main features of the KAMIStudio environment can be used for semi-automatic curation of large corpora of cellular signalling knowledge and for dynamic study of the modelled systems.

Round table discussion

The general discussion highlighted four important aspects, namely (1) the need to provide annotated models that would include textual annotations, bibliographic references and crosslinks to knowledge resources through the use of common identifiers, (2) the importance of creating interfaces for automatic integration of annotations by leveraging the wealth of curated interactions in dedicated databases, (3) the utility of agreeing on best practices, use of standards and on the minimum information required to ensure model reproducibility and reusability, and lastly (4) the use of common repositories for logical models that would foster interactions and facilitate exchanges between scientists interested in reusing models. The need to encourage novel publications with logical models to be systematically submitted to one of the model repositories was also discussed, as this would increase visibility, ease reproducibility, and promote reusability of logical models.

Roadmap to best practices for the Curation and Annotation of Logical Models (CALM) in biology

Based on these discussions, four interdependent milestones were identified for the roadmap to curation and annotation of logical models in biology (Figure 2):

a) The first milestone concerns the **reproducibility of the analyses of discrete models**. The use of common, standardised formats (e.g., SBML packages *qual*, *layout*, *render*, etc.) would greatly facilitate the interoperability between different tools and the development of integrative pipelines. For example, the CoLoMoTo notebook could be expanded to include more tools, offering a flexible way of performing dynamical analyses in a fully transparent and reproducible manner. To achieve this goal, the logical modelling community aims to work close with the communities developing standards, such as SBML, the Simulation Experiment Description Markup Language (SED-ML) and Computational Modelling in Biology Network (COMBINE) to contribute to community efforts and make sure that the standards developed are in line with the specificities of the logical formalism.

b) The second milestone concerns the **minimum information for annotating a model, and also new mechanisms to encode such information in SBML-*qual*.** The information should be stored in human and machine-readable form, for example, by using Resource Description Framework (RDF) tags [63]. SBML format also provides the possibility to associate Systems Biology Ontology (SBO) terms outside of RDFs; however, unified storage of all model annotations in RDF could provide a simple, yet an efficient standard way of annotating logical models. Supported by larger computational modelling communities (e.g., COMBINE), RDF is considered the *de facto* standard for encoding annotations [64]. The community should discuss and agree on the best way of integrating annotations in SBML-*qual*, notably which tags and which SBML elements to use, while also leveraging the experience of the SBML community and BioModels curation practices. Notably, the SBML specification documents [7]

already propose a systematic way of annotation that can be adapted to logical models. Additionally, the logical modelling community should define specific needs that are not covered yet by existing standards (i.e., MIRIAM identifiers and BioModels.net qualifiers [65]) and propose feasible solutions. The minimum information for annotation could be proposed as a prerequisite for publishing a logical model in peer-reviewed journals. Table 2 lists suggested minimum qualifiers that could be used in order to annotate a model, in line with MIRIAM and BioModels suggestions. Furthermore, to aid model developers and curators, new tools need to be developed to aid the enrichment of models with as many relevant and useful annotations as possible. The metadata information for one of the three curated logical models currently available in BioModels and the corresponding code block of the xml file are exemplified in Figure 3. While the logical modelling community has made progress towards identifying important aspects of annotations, much work remains to be done. For example, the community is currently discussing the appropriate "depth" of annotations for each logical function. For example, does each variable and operator between variables in a logical function need to be annotated (such as currently available in Cell Collective)? While this level of annotations can add to the work-load of the modeller/curator, one might argue that providing citable experimental evidence for such aspects for the regulatory mechanism of each component will only increase the transparency of the model. The qual:transition component in the SBML model could be proposed for the annotation of causal interaction, however, this choice (already employed by some tools i.e., CaSQ, Cell Collective) raises issues concerning the cases where a more precise annotation would be needed.

c) The third milestone refers to the **collaboration between modellers and curators to bridge the gap between storing information and reusing information.** Automated

procedures for model annotation and enrichment could further help to maintain models up to date. Keeping track of literature information used to derive logical formulae can further foster model accuracy and enhance reusability. To make steps forward, the logical modelling community aims to work closely with biocurators and knowledge platform developers to identify best practices. An obvious way would be to agree on the use of common and well-established identifiers like UniProt, GO, HGCN, SBO that would allow unambiguous identification of a model component with simultaneous access to the knowledge resource through crosslinks. This direct linking of model annotations to curated knowledge sources via standard identifiers could help significantly in establishing quality control checks regarding annotation and biological content.

d) The last milestone concerns **fully leveraging available model repositories**. Several logical model repositories exist, including Cell Collective, GINsim and PyBoolNet [66]. The Cell Collective provides models in several formats, including SBML *qual*. The GINsim's model repository provides models in the GINML format, which can be converted to SBML *qual* (and other formats) using GINsim and BioLQM. Simultaneously, BioModels is one of the largest repositories of mathematical, SBML-encoded models. However, it has been traditionally focused on models described with other mathematical frameworks, and lacks processes to curate logical models. Indeed, the logical modelling community has started to work closely with BioModels team to set up best practices and model quality checks that will be applicable to logical models. The aim is to create a dedicated collection of logical models within BioModels, which would provide an additional resource with curated logical models. In Box 1, we show a curated logical model stored in BioModels (BIOMD0000000593) annotated as a sample case.

The logical modelling community should also decide if the suggestions of the COMBINE community, as stated in Neal et al. [64], regarding the storage of annotations in a separate file could be adopted. While this would allow for more flexibility in terms of knowledge resources' choices for model annotation, i.e., one model file with several annotation files with different sources, it would add the extra burden of file synchronization. However, dissociating model from model annotation could be in line with the approaches and methodologies presented in the first session of the meeting regarding the separation of the biocuration from the model layout and refinement. An additional point to consider is the simulation settings and their specifications through an established standard such as SED-ML [67,68], which will likely require some adaptation to suit logical model simulations. In this respect, the COMBINE Archive format could offer a possible solution, as it provides a standardised way to bundle this type of files together [64].

Key Points

- The identified milestones will help the community of logical modelling to coordinate efforts for reproducible research.
- Standards for minimum curation will help unify formats and annotations, in an effort to provide models of better accuracy and quality.
- Transparency in curation and standardised annotations will enhance model reusability.
- Format harmonisation will facilitate interoperability and integration of existing tools in seamless pipelines.
- Collaboration between modellers and curators will foster model enrichment and updating, taking advantage of the wealth of information stored in databases and knowledge bases.

- The use of a common repository will reinforce quality protocols and checks for models, which could even be used prior to publication.

Outcomes and Outlook

The [BC]2 workshop on annotation and curation of logical models in biology brought together people from different communities, such as biocurators, modellers, methodology and software developers. The round table discussion clarified common objectives together with milestones on the roadmap to best practices. Presentations and discussions highlighted efforts and resources that can be used for enhancing reproducibility and model contextualisation. The authors have started to form working groups and will continue to foster communication and exchanges first among the logical modelling community and also by reaching out to other communities with similar interests, to attain these collective goals.

The complete list of abstracts can be found in the supplementary material *Abstract_Booklet*.

Authors' information

Anna Niarakis is an Associate Professor at Univ Evry, University of Paris-Saclay. Her research focuses on the application of Computational Systems Biology approaches in human diseases, including the construction of disease maps, tool development for model inference, network integration and dynamical modelling.

Martin Kuiper is Professor in Systems Biology at the Department of Biology of the Norwegian University of Science and Technology. His group develops enabling technology for data analysis and knowledge management, among others for the construction of logical models for systems medicine.

Marek Ostaszewski is a scientist and a project manager at the LCSB, working on knowledge management and visualization in systems biomedicine, in particular in Parkinson's disease, including clinical research.

Rahuman S Malik Sheriff is a project leader at the European Bioinformatics Institute (EMBL-EBI). He leads the development of BioModels infrastructure as well as curation and annotation of computational models in the standard formats. He is one of the members of the SBML Editorial Board. His interest includes building tools and resources for system biology modelling.

Cristina Casals-Casas is a biocurator in the Swiss-Prot group at SIB. Her main responsibilities include the literature-based expert curation of mammalian proteins for UniProtKB and GO knowledge bases. She is also involved in research projects that require expert biocuration.

Denis Thieffry is currently Professor of Systems Biology at the Ecole Normale Supérieure (ENS), Paris, France, and part-time Invited Research Professor at the Cancer Science Institute of the National University of Singapore.

Tom C. Freeman is Professor holding the chair of Systems Immunology at the University of Edinburgh. He has a broad range of interests ranging from immune cell differentiation, transcriptomics, data visualisation, and pathway modelling and simulation.

Paul Thomas is an Associate Professor at the University of Southern California. His research is in the area of computational analysis of genomic data, with an emphasis on gene function and evolution. In addition to founding and continuing development on the PANTHER project, Dr. Thomas is a director of the Gene Ontology Consortium, one of the largest and best-known bioinformatics projects in the world.

Vasundra Touré is a PhD student at the Norwegian University of Science and Technology. She is interested in the knowledge management and data extraction of molecular causal statements.

Vincent Noël is a postdoctoral fellow currently working in Institut Curie. His research interests concern modelling of biological systems, from the construction and the analysis of models to the optimization of their simulations.

Julio Saez-Rodriguez is Professor of Medical Bioinformatics and Data Analysis at the Faculty of Medicine of Heidelberg University, director of the Institute for Computational Biomedicine, group leader of the EMBL-Heidelberg University Molecular Medicine Partnership Unit and a co-director of the DREAM challenges. His research focuses on computational methods to understand and treat the deregulation of cellular networks in disease.

Aurélien Naldi is a postdoctoral fellow currently working in the Ecole Normale Supérieure (ENS). His interests cover the integration of advanced computational methods into software tools for the analysis of complex biological models.

Eugenia Oshurko is a PhD student at ENS Lyon. Her research interests include graph rewriting, graph-based databases, and knowledge representation and their application to the curation of biological knowledge relevant to building executable rule-based models.

Ioannis Xenarios is an Associate Professor of Computational Biology at the CHUV/UNIL/LICR. His research aims to model the tumor microenvironment of T cells infiltrates (TIL) and monocytes. He develops both curated and data driven based model to characterize and provide experimentally testable hypothesis using drugs and immunotherapies.

Sylvain Soliman is a researcher at Inria, Saclay. His research interests focus around Computational Biology and Theoretical Computer Science. He has been one of the main developers and maintainers of the BIOCHAM modelling and analysis platform for more than ten years.

Claudine Chaouiya is an Associate Professor at I2M, Aix-Marseille University. Her research focuses on developing appropriate methods for the analysis of large logical models of regulatory and signalling networks. Methodological advancements and tool development are motivated and further validated through modelling projects revolving around cell fate decisions, in single and multi-cellular contexts.

Tomas Helikar is an Associate Professor in the Department of Biochemistry at the University of Nebraska-Lincoln. His research focuses on the use of integrative multi-approach modelling pipelines for dynamical analysis of biological networks. More precisely, his studies focus on understanding how aberrant changes in biological networks result in disease so that we could strategically develop more effective therapies.

Laurence Calzone is a researcher in Institut Curie. Her research activities revolve around building knowledge maps, constructing logical models, and developing tools for improving and optimizing the qualitative models and their predictions.

Acknowledgements

The authors would like to thank the SysMod community for its support, the [BC]2 organisers for supporting and hosting the meeting/workshop in Basel and Krishna Kumar Tiwari, Scientific Database Curator in BioModels, for his help in producing the curated example depicted in Box 1.

Funding

Not applicable

Competing Interests

The authors declare no competing interests regarding the content of this manuscript

References

1. Allison DB, Shiffrin RM, Stodden V. Reproducibility of research: Issues and proposed remedies. Proc. Natl. Acad. Sci. 2018; 115:2561–2562
2. Naldi A, Hernandez C, Abou-Jaoudé W, et al. Logical Modeling and Analysis of Cellular Regulatory Networks With GINsim 3.0. Front. Physiol. 2018; 9:646
3. Helikar T, Kowal B, McClenathan S, et al. The Cell Collective: Toward an open and collaborative approach to systems biology. BMC Syst. Biol. 2012; 6:96
4. Helikar T, Kowal B, Rogers JA. A cell simulator platform: the cell collective. Clin. Pharmacol. Ther. 2013; 93:393–395
5. Malik-Sheriff RS, Glont M, Nguyen TVN, et al. BioModels—15 years of sharing computational models in life science. Nucleic Acids Res. 2019;
6. Hucka M, Finney A, Sauro HM, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 2003; 19:524–531
7. Chaouiya C, Keating SM, Berenguier D, et al. SBML Level 3 package: Qualitative Models, Version 1, Release 1. J. Integr. Bioinforma. 2015; 12:2
8. Chaouiya C, Bérenguier D, Keating SM, et al. SBML qualitative models: a model representation format and infrastructure to foster interactions between qualitative modelling formalisms and tools. BMC Syst. Biol. 2013; 7:135
9. Naldi A, Monteiro PT, Müssel C, et al. Cooperative development of logical modelling

standards and tools with CoLoMoTo. Bioinformatics. 2015; 31:7; 1154–1159
10. SysMod. SysMod Community of Special Interest. http://sysmod.info
11. BASEL LIFE 2019, 9–12 September 2019, Congress Center Basel, Switzerland. https://www.basellife.org/2019.html
12. BaselLife 2019 Workshops and tutorials. https://www.basellife.org/2019/basel-life-structure/bc2/programme/workshops-and-tutorials.html#c68317
13. BioKB. https://biokb.lcsb.uni.lu/
14. Gawron P, Ostaszewski M, Satagopam V, et al. MINERVA-a platform for visualization and curation of molecular interaction networks. NPJ Syst. Biol. Appl. 2016; 2:16020
15. Bateman A, Martin MJ, O'Donovan C, et al. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017; 45:D158–D169
16. Perfetto L, Briganti L, Calderone A, et al. SIGNOR: a database of causal relationships between biological entities. Nucleic Acids Res. 2016; 44:D548–D554
17. Braschi B, Denny P, Gray K, et al. Genenames.org: the HGNC and VGNC resources in 2019. Nucleic Acids Res. 2019; 47:D786–D792
18. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019; 47:D330–D338
19. Fabregat A, Jupe S, Matthews L, et al. The Reactome Pathway Knowledgebase. Nucleic Acids Res. 2018; 46:D649–D655
20. The Druglogics Initiative. https://www.druglogics.eu/
21. vsmjs. https://github.com/vsmjs. GitHub 2019
22. SciCura. http://demo.scicura.org/
23. MI2CAST.https://github.com/MI2CAST
24. Biryukov M, Groues V, Satagopam V. BioKB - Text Mining and Semantic Technologies for the Biomedical Content Discovery. University of Luxembourg. 2019; https://pdfs.semanticscholar.org/a286/dabfaa6356ac0634a9ace69e0d6ac289dfb4.pdf
25. Hoksza D, Gawron P, Ostaszewski M, et al. MINERVA API and plugins: opening molecular network analysis and visualization to the community. Bioinformatics 2019; 35:4496–4498
26. Hoksza D, Gawron P, Ostaszewski M, et al. Closing the gap between formats for storing layout information in systems biology. Brief. Bioinform. 2019;
27. Kitano H, Funahashi A, Matsuoka Y, et al. Using process diagrams for the graphical representation of biological networks. Nat. Biotechnol. 2005; 23:961–966
28. Newt Editor. http://newteditor.org/index.html
29. Glont M, Nguyen TVN, Graesslin M, et al. BioModels: expanding horizons to include more modelling approaches and formats. Nucleic Acids Res. 2018; 46:D1248–D1253
30. Le Novère N, Finney A, Hucka M, et al. Minimum information requested in the annotation of biochemical models (MIRIAM). Nat. Biotechnol. 2005; 23:1509–1515
31. Yates AD, Achuthan P, Akanni W, et al. Ensembl 2020. Nucleic Acids Res. 2019;
32. Federhen S. The NCBI Taxonomy database. Nucleic Acids Res. 2012; 40:D136–D143
33. Degtyarenko K, de Matos P, Ennis M, et al. ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res. 2008; 36:D344–D350
34. Mathematical Modelling Ontology. https://www.ebi.ac.uk/ols/ontologies/mamo
35. Systems Biology Ontology. http://www.ebi.ac.uk/sbo/main/
36. BRENDA tissue / enzyme source. https://www.ebi.ac.uk/ols/ontologies/bto
37. Bekkar A, Estreicher A, Niknejad A, et al. Expert curation for building network-based dynamical models: a case study on atherosclerotic plaque formation. Database J. Biol. Databases Curation 2018; 2018:
38. Naldi A, Hernandez C, Levy N, et al. The CoLoMoTo Interactive Notebook: Accessible and Reproducible Computational Analyses for Qualitative Biological Networks. Front.

Physiol. 2018; 9:
39. Freeman TC, Raza S, Theocharidis A, et al. The mEPN scheme: an intuitive and flexible graphical system for rendering biological pathways. BMC Syst. Biol. 2010; 4:65
40. Ivy JR, Shih B, Hogenesch JB, et al. A detailed graphical and computational model of the mammalian renal circadian clock. bioRxiv 2019; 795906
41. Livigni A, O'Hara L, Polak ME, et al. Petri Net-Based Graphical and Computational Modelling of Biological Systems. bioRxiv 2016; 047043
42. Aghamiri, S.S, Singh, V, Naldi A, et al. Automated inference of Boolean models from molecular interaction maps using CaSQ. Bioinformatics *Under review*
43. Singh V, Ostaszewski M, Kalliolias GD, et al. Computational Systems Biology Approach for the Study of Rheumatoid Arthritis: From a Molecular Map to a Dynamical Model. Genomics Comput. Biol. 2018; 4:1
44. Singh V, Kalliolias GD, Ostaszewski M, et al. RA-map: Building a state-of-the-art interactive knowledge base for rheumatoid arthritis. Database *accepted*;
45. Helikar T, Cutucache CE, Dahlquist LM, et al. Integrating interactive computational modeling in biology curricula. PLoS Comput. Biol. 2015; 11:e1004131
46. Kowal BM, Schreier TR, Dauer JT, et al. Programmatic access to logical models in the Cell Collective modeling environment via a REST API. Biosystems 2016; 139:12–16
47. Stoll G, Caron B, Viara E, et al. MaBoSS 2.0: an environment for stochastic Boolean modeling. Bioinformatics 2017; 33:2226–2228
48. Letort G, Montagud A, Stoll G, et al. PhysiBoSS: a multi-scale agent-based modelling framework integrating physical dimension and cell signalling. Bioinformatics 2019; 35:1188–1196
49. pyMaBoSS — pyMaBoSS 0.3.2 documentation. https://pymaboss.readthedocs.io/en/latest/
50. Orchard S, Kerrien S, Abbani S, et al. Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. Nat. Methods 2012; 9:345–350
51. Taylor CF, Hermjakob H, Julian RK, et al. The work of the Human Proteome Organisation's Proteomics Standards Initiative (HUPO PSI). Omics J. Integr. Biol. 2006; 10:145–151
52. causalBuilder. https://vtoure.github.io/causalBuilder//causalBuilder/. causalBuilder
53. Terfve CDA, Wilkes EH, Casado P, et al. Large-scale models of signal propagation in human cells derived from discovery phosphoproteomic data. Nat. Commun. 2015; 6:1–11
54. Liu A, Trairatphisan P, Gjerga E, et al. From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL. Npj Syst. Biol. Appl. 2019; 5:1–10
55. Türei D, Korcsmáros T, Saez-Rodriguez J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat. Methods 2016; 13:966–967
56. Ceccarelli F, Turei D, Gabor A, et al. Bringing data from curated pathway resources to Cytoscape with OmniPath. Bioinformatics 2019; btz968
57. Naldi A. BioLQM: A Java Toolkit for the Manipulation and Conversion of Logical Qualitative Models of Biological Networks. Front. Physiol. 2018; 9
58. Paulevé L. Pint: A Static Analyzer for Transient Dynamics of Qualitative Networks with IPython Interface. 2017; https://hal.archives-ouvertes.fr/hal-01589248
59. KAMIStudio. http://kamistudio.ens-lyon.fr/about
60. Boutillier P, Maasha M, Li X, et al. The Kappa platform for rule-based modeling. Bioinformatics 2018; 34:i583–i592
61. Faeder JR, Blinov ML, Hlavacek WS. Rule-Based Modeling of Biochemical Systems with BioNetGen. Syst. Biol. 2009; 113–167
62. Kappa-Dev/KAMI. https://github.com/Kappa-Dev/KAMI

63. Novère NL, Finney A. A simple scheme for annotating SBML with references to controlled vocabularies and database entries. 2005; 13
64. Neal ML, König M, Nickerson D, et al. Harmonizing semantic annotations for computational models in biology. Brief. Bioinform. 2019; 20:540–550
65. BioModels.net Qualifiers | COMBINE. http://co.mbine.org/specifications/qualifiers
66. Klarner H, Streck A, Siebert H. PyBoolNet: a python package for the generation, analysis and visualization of boolean networks. Bioinforma. Oxf. Engl. 2017; 33:770–772
67. Bergmann FT, Cooper J, König M, et al. Simulation Experiment Description Markup Language (SED-ML) Level 1 Version 3 (L1V3). J. Integr. Bioinforma. 2018; 15
68. Waltemath D, Adams R, Bergmann FT, et al. Reproducible computational biology experiments with SED-ML - The Simulation Experiment Description Markup Language. BMC Syst. Biol. 2011; 5:198

Figure Legends

Figure 1. Four main thematic axes of the presentations and the round table discussion of the meeting. Biocuration platforms, available model repositories, tool development and integrative methodologies were the main subjects of the meeting. All presentations highlighted the need for standards in model annotation and curation.

Figure 2. Roadmap to Curation and Annotation of Logical Models in Biology. Four milestones were identified as key steps in the roadmap to best practices for logical models annotation and curation: integrated pipelines for reproducible research, standards for SBML qual annotations, automation of models enrichment and the use of a common repository.

Figure 3. A logical model in Biomodels database. Metadata information for the curated logical model in BioModels database (upper panel) and the corresponding block code (lower panel).

Box 1. An example of annotating a logical model using RDFs. BioModels propose a two level annotation, model and model component. Model components are in turn annotated in two

levels: nodes and arcs/ interactions A color code is used to highlight the different code blocks that refer to each level of annotation. Code blocks are excerpts from a syntactically valid SBML *qual* file.

Tables

Table 1: Summary of different topics and presentations.

Table 2: Suggestion of minimum qualifiers for the annotation of logical models. The hasState qualifier could be added to account for a node's state (qualitative levels).

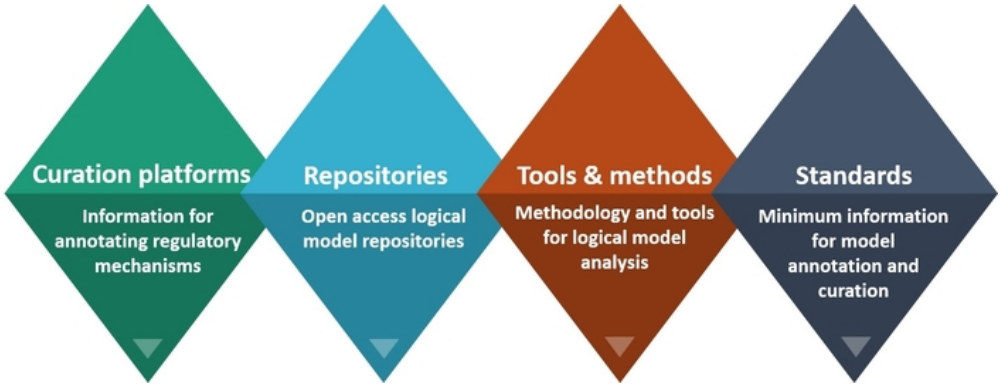


Figure 1. Four main thematic axes of the presentations and the round table discussion of the meeting. Biocuration platforms, available model repositories, tool development and integrative methodologies were the main subjects of the meeting. All presentations highlighted the need for standards in model annotation and curation.

59x22mm (300 x 300 DPI)

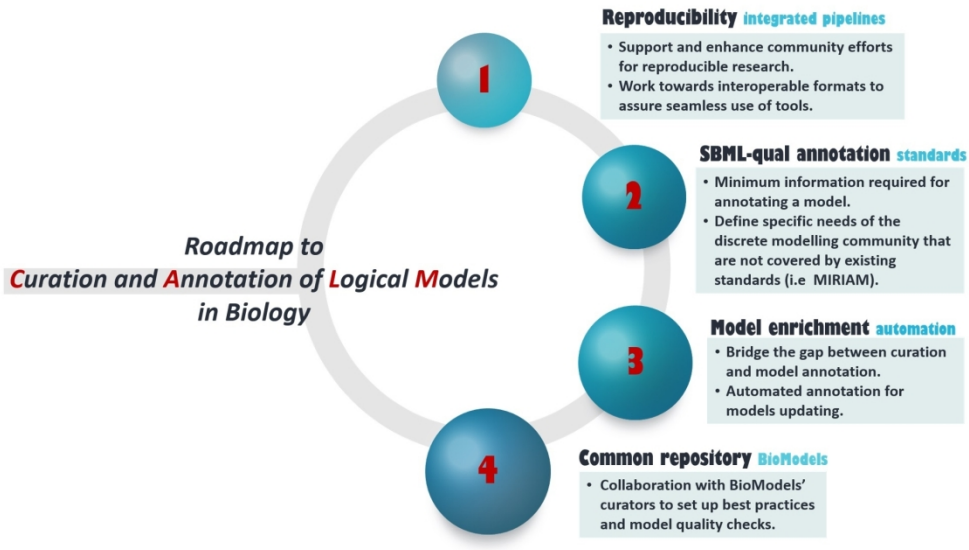


Figure 2. Roadmap to Curation and Annotation of Logical Models in Biology. Four milestones were identified as key steps in the roadmap to best practices for logical models annotation and curation: integrated pipelines for reproducible research, standards for SBML qual annotations, automation of models enrichment and the use of a common repository.

84x47mm (600 x 600 DPI)

Figure 3. A logical model in Biomodels database. Metadata information for the curated logical model in BioModels database (upper panel) and the corresponding block code (lower panel).

32x39mm (300 x 300 DPI)

Use case of a curated and annotated logic model (BIOMD0000000593) in BioModels.

- Cross-references to well-established ontologies like GO, UNIPROT, NICT, SBO etc. are added to NODES (SPECIES) and CAUSAL-INTERACTIONS using RDF.
- Use of COMBINE qualifiers http://co.mbine.org/standards/qualifiers where possible.

Annotations are added at two levels:

1. **Model Level annotation**

```
<bqmodel:isDerivedFrom>
    <rdf:Bag>
        <rdf:li rdf:resource="http://identifiers.org/pubmed/12871957"/>
        <rdf:li rdf:resource="http://identifiers.org/pubmed/16314431"/>
    </rdf:Bag>
</bqmodel:isDerivedFrom>
```

2. **Model component level annotation:**
    a. **NODE (SPECIES)**

```
<qual:qualitativeSpecies metaid="species11" qual:compartment="default"
qual:constant="false" qual:id="IL21" qual:maxLevel="1" qual:name="IL21">
    <annotation>
        <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:dc="http://purl.org/dc/elements/1.1/"
        xmlns:dcterms="http://purl.org/dc/terms/"
        xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
            <rdf:Description rdf:about="#species11">
                <bqbiol:is>
                    <rdf:Bag>
                        <rdf:li rdf:resource="http://identifiers.org/uniprot/Q9HBE4" />
                    </rdf:Bag>
                </bqbiol:is>
            </rdf:Description>
        </rdf:RDF>
    </annotation>
</qual:qualitativeSpecies>
```

Box 1. An example of annotating a logical model using RDFs. BioModels propose a two level annotation, model and model component. Model components are in turn annotated in two levels: nodes and arcs/ interactions A color code is used to highlight the different code blocks that refer to each level of annotation. Code blocks are excerpts from a syntactically valid SBML qual file.

32x30mm (300 x 300 DPI)

2. Model component level annotation:
b. CAUSAL INTERACTIONS

```
<qual:transition metaid="rxn01" qual:id="tr_TBET" qualname="Interactions targeting TBET">
    <qual:listOfInputs>
        <qual:input metaid="rxn01_input01" qual:qualitativeSpecies="IL4"
        qual:transitionEffect="none"/>
    </qual:listOfInputs>
    <qual:listOfOutputs>
        <qual:output metaid="rxn01_output01" qual:qualitativeSpecies="TBET"
        qual:transitionEffect="assignmentLevel"/>
    </qual:listOfOutputs>
    <qual:listOfFunctionTerms>
        <qual:functionTerm metaid="rxn01_function01" qual:resultLevel="1">
            <math xmlns="http://www.w3.org/1998/Math/MathML">
                <apply>
                    <not/>
                    <apply>
                        <or
                            />
                        <apply>
                            <eq/>
                            <ci> IL4 </ci>
                            <cn type="integer"> 1 </cn>
                        </apply>
                    </apply>
                </apply>
            </math>

            <annotation>
                <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
                xmlns:dc="http://purl.org/dc/elements/1.1/"
                xmlns:dcterms="http://purl.org/dc/terms/"
                xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
                    <rdf:Description rdf:about="#_rxnannotation1">
                        <bqbiol:hasProperty>
                            <rdf:Bag>
                                <rdf:li
                                rdf:resource="http://identifiers.org/SBO/SBO:0000169"/>
                            </rdf:Bag>
                        </bqbiol:hasProperty>
                    </rdf:Description>
                </rdf:RDF>
            </annotation>
        </qual:functionTerm>
        <qual:defaultTerm metaid="_defaultlevel" qual:resultLevel="0">
        </qual:defaultTerm>
    </qual:listOfFunctionTerms>
</qual:transition>
```

Box 1. An example of annotating a logical model using RDFs. BioModels propose a two level annotation, model and model component. Model components are in turn annotated in two levels: nodes and arcs/ interactions A color code is used to highlight the different code blocks that refer to each level of annotation. Code blocks are excerpts from a syntactically valid SBML qual file.

31x34mm (300 x 300 DPI)

Table 1: Summary of different topics and presentations.

Workshop sessions and Chairs	Presentations and speakers
Model curation and annotation and available repositories ***Chairs:*** ***Anna Niarakis and Denis Thieffry***	• **Martin Kuiper:** Towards a curation platform for causal interaction statements. • **Marek Ostaszewski:** BioKB and MINERVA: a workflow for curation and quick prototyping of annotated knowledge repositories • **Rahuman S Malik Sheriff:** Curation and annotation of models in BioModels repository promotes reproducibility and reusability • **Cristina Casals:** SysVasc Prior Knowledge Network: An example of biocuration for Boolean modelling
Community standards development and interoperability/reusability ***Chairs:*** ***Marek Ostaszewski and Laurence Calzone***	• **Denis Thieffry:** Computational verification of large logical models - application to the prediction of T cell response to checkpoint inhibitors • **Tom Freeman:** A graphical and computational model of the renal mammalian circadian clock • **Paul Thomas:** Gene Ontology Causal Activity Modeling • **Anna Niarakis:** Automated inference of annotated Boolean models from molecular interaction maps using CaSQ
Tools (I) ***Chair:*** ***Julio Saez Rodriguez***	• **Tomas Helikar:** Cell Collective modelling platform • **Gaultier Stoll and Vincent Noel:** MaBoSS ecosystem • **Vasundra Touré:** The Minimum Information about a Molecular Interaction Causal Statement (MI2CAST): a guideline for the annotation of molecular causal interactions
Tools (II) ***Chair:*** ***Tomas Helikar***	• **Julio Saez Rodriguez:** CellNOpt • **Aurélien Naldi:** The CoLoMoTo Interactive Notebook: Accessible and Reproducible Computational Analyses for Qualitative Biological Networks • **Eugenia Oshurko:** KAMIStudio

Table 2: Suggestion of minimum qualifiers for the annotation of logical models. The hasState qualifier could be added to account for a node's state (qualitative levels).

Model annotation levels	Minimum Qualifiers	Examples of knowledge sources stored in RDFs
Model	**Model Qualifiers: bqmodel** **is, identity** This qualifier might be used to link an encoded model to a database of models. **isDescribedBy, description** This relation might be used to link a model to the literature that describes it. **hasTaxon, taxon** This qualifier might be used to indicate taxonomy/ organism (i.e: human, plant, animal). **isVersionOf, version** This qualifier can be used to link a model to the Gene Ontology terms regarding the biological function described. **hasProperty, property** This relation could be used to indicate mathematical formalism. **isDerivedFrom, origin** This relation may be used to express a refinement or adaptation in usage for a previously described model	PMID, BioModels ID, doi, CC ID, GINsim ID, GO
Qualitative Species	**Biology Qualifiers: bqbiol** **is, identity** This relation might be used to link a biological entity to its exact counterpart in a database. **isDescribedBy, description** This relation should be used to link a species to the literature that describes the role of that species or its presence in the system of interest. **hasVersion, version** This relation may be used to represent an isoform or modified form of a biological entity. **hasState, state** This relation could be used to describe the state of a biological entity.	GO, UniProt, HGCN, PMID
Causal interactions/transitions	**Biology Qualifiers: bqbiol** **hasProperty, property** This relation might be used when a biological entity exhibits a	KEGG, REACTOME, PMID

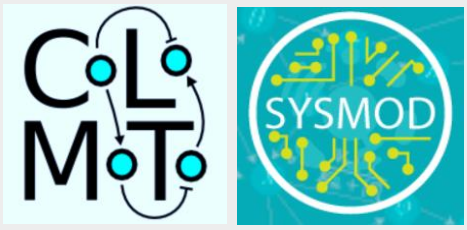
	certain enzymatic activity or exerts a specific function. **isDescribedBy, description** This relation should be used, for instance, to link a reaction to the literature that describes it.	

WS2: Annotation and curation of computational models in biology

September 9, 2019 – University of Basel

Kollegienhaus, Petersplatz 1, CH-4001 Basel.

ABSTRACT BOOKLET



Room

Hörsaal 116

Organisers

All organisers are members of the CoLoMoTo/SysMod communities, with experience in workshop organisation.

- **Anna Niarakis**, Univ Evry, University of Paris-Saclay; FR
- **Tomas Helikar**, University of Nebraska; USA
- **Laurence Calzone**, Institut Curie/U900, INSERM/Mines ParisTech; FR

Overview

The fast accumulation of biological data calls for more systematic approaches for their integration, analysis and exploitation. The generation of novel, relevant hypotheses from this enormous quantity of data remains challenging. Logical models have long been used to answer a variety of questions regarding the dynamical behaviours of regulatory networks. As the number of published logical models increases, there is a pressing need for proper model annotation, referencing and curation in community-supported and standardised formats. In this context, organised by members of the *Consortium for Logical Models and Tools* (CoLoMoTo – http://colomoto.org) and of the *Computational Modeling of Biological Systems Community of Special Interest (COSI) of the International Society for Computational Biology (ISCB) (*SysMod - https://sysmod.info/), this workshop aims to review and connect different ongoing projects, bringing together people from different communities involved in modelling and annotation of molecular biological entities, interactions, pathways and models.

Invited speakers

- **Martin Kuiper**, Norwegian University of Science and Technology, NO**,**
- **Denis Thieffry**, IBENS, Paris, FR
- **Rahuman S. Malik Sheriff**, Project Leader (BioModels), EMBL-EBI, London, UK
- **Cristina Casals**, UniProtKB/ Swiss-Prot Biocurator, SIB, Geneva, CH
- **Paul Thomas** – GO, (LEGO/Noctua), USC, USA

Session 1 – Data/model curation/annotation and available repositories

Martin Kuiper: Towards a curation platform for causal interaction statements.

The DrugLogics group (https://www.druglogics.eu/) at NTNU develops logical modelling of cancer cell systems to predict cell fate effects of targeted drugs and drug combinations. We follow two paths towards assembling models for computer simulation: manual curation, obtaining logical model components and regulatory relationships from knowledge bases and the literature, and automated model topology building from causal interaction statements: basic representations of regulatory interactions between two biological entities that are extracted from cell signaling databases like Signor and Reactome. These causal interactions are automatically incorporated into a logical model based on 1) containing a drug target; 2) linking to model 'output' nodes (pro-survival and anti-survival), and 3) providing connectivity between network nodes, until a self-contained model is obtained. The accuracy of these automatically built models depends, among others, on the richness of the available set of causal statements. We have therefore set out to implement a novel curation interface called VSM (http://scicura.org/vsm/intro.html) for the curation of causal statements. VSM offers a very versatile curation interface to annotate information from molecular biology or in fact any other domain, and we have used a VSM template interface to annotate the full collection of experimentally analysed DNA binding Transcription Factors for human, mouse and rat. Transcription factor – target gene interactions represent a valuable class of causal interactions for modelling. Extending on this, within the COLOSYS project (https://www.colosys.org/) we committed to deliver a standard and guidelines for general molecular causal interaction curation: Minimum Information about a Molecular Interaction Causal Statement, MI2CAST). We have now started to embed this standard into the VSM curation interface. A detailed introduction into MI2CAST and how it is enabled by VSM templates will be given by Vasundra Touré.

Marek Ostaszewski: BioKB and MINERVA: a workflow for curation and quick prototyping of annotated knowledge repositories

Developing new models in systems biology is challenging because of the laborious process of biocuration needed to explore, standardize and encode the abundance of information to available in literature. This work becomes even more difficult when a diagram representation of the model is needed, or, in case of kinetic model, when the model needs to be parametrized. The construction of diagrammatic models, realized with the help of dedicated software like CellDesigner, Newt or Krayon, requires performing two tasks at the same time: i) extracting the entities, their relationships and annotations from the literature and ii) providing a consistent graphical representation of these entities and relationships. In effect, diagram editors do offer annotation functionalities, but they are limited and cumbersome, especially when multiple annotations or source sentences are needed. In the talk, we will propose to separate these two tasks into a workflow, with biocuration of facts from literature being the first step, and the diagram layout and model refinement being the

second step. The first step is realized with the help of the BioKB (biokb.lcsb.uni.lu), a platform for exploring text mining data. We introduced a biocurator interface to BioKB, allowing to refine machine-identified interactions into "facts" – human-curated relationships, annotated with sentences, literature and recognized identifiers. BioKB "facts" can be reviewed by other platform users and are version-tracked. Curators can use filters to choose their starting point in the text mining repository, combine multiple interactions into a single "fact", or introduce their own sentences. The second step of the curation workflow is the upload of the curated content into a diagram editor for layout and refinement. Here, we will present the capabilities of MINERVA platform to upload a layout-less model and convert it into an editable diagram using dedicated API calls. We will discuss the capabilities and shortcomings of MINERVA in cross-format conversion between CellDesigner SBML, layout and render SBML and SBGN-ML. In summary, we will propose and discuss the improvement of a curation workflow for diagrammatic models from a process centered around a diagram editor to a workflow separating the curation of annotated literature facts from constructing a comprehensive diagram based on these facts. With this we hope to contribute to the process of creating high quality models in systems biology.

Sheriff Malik: Curation and annotation of models in BioModels repository promotes reproducibility and reusability

Models of cell signalling, metabolic and gene regulatory networks have been shown to divulge mechanistic insight into cellular regulation. To provide a platform to support universal sharing, easy accessibility and model reproducibility, BioModels (www.ebi.ac.uk/biomodels/), a repository for mathematical models was established in 2005 (Chelliah *et al.* 2015; Glont *et al.* 2018). Models submitted to BioModels are curated to verify the computational representation of the biological process and to reproduce the simulation results in the reference publication. Following MIRIAM guidelines, the curated models are encoded in the standard SBML format and semantically enriched with controlled vocabularies (Le Novère *et al*. 2005). Model entities are cross-referenced to several data resources (such as UniProt, Ensembl gene, taxonomy, etc.,) as well as ontologies (such as Gene Ontology, ChEBI, Mathematical Modelling Ontology, Systems Biology Ontology, Brenda Tissue Ontology, etc.,). These annotations allow unambiguous identification of model components and processes as well as make models searchable. With gradual curation efforts, BioModels currently hosts 800 curated models, becoming the world's largest repository of curated models. BioModels has also emerged as third most used data resource after PubMed and Google Scholar among the scientists who use modelling in their research (Stanford *et al.* 2015; Szigeti *et al.* 2018). Thus, the model curation and annotation in BioModels is instrumental in providing reproducible and semantically enriched models in standard formats that significantly benefits modellers.

Cristina Casals: SysVasc Prior Knowledge Network: An example of biocuration for Boolean modelling

New types of computational analyses, such as network-based dynamical models, take advantage of the enormous amounts of data that biologists today are able to generate at relatively low cost. However, expert curation remains critical to transform available data—including publications—into computable knowledge. In this work, we demonstrate how expert curation plays a direct role in research, by supporting the use of network-based dynamical models to study a specific biological process. This curation effort is focused on the regulatory interactions between biological entities, such as genes or proteins and compounds, which may interact with each other in a complex manner, including regulatory complexes and conditional dependencies between co-regulators. This critical information has to be captured and encoded in a computable manner, which is currently far beyond the current capabilities of automatically constructed network. As a case study, we report here the prior knowledge network constructed by the sysVASC consortium to model the biological events leading to the formation of atherosclerotic plaques, during the onset of cardiovascular disease. We also discuss some specific examples to illustrate the main pitfalls and added value provided by the expert curation during this endeavor.

Session 2 – Community standards development and interoperability / reusability

Denis Thieffry: Computational verification of large logical models - application to the prediction of T cell response to checkpoint inhibitors

At the crossroad between biology and computational modelling, systems biology has proved to be an important ally to gain a mechanistic understanding of biological systems. But as our knowledge accumulates, the size and complexity of mathematical models increase, calling for the development of efficient dynamical analysis methods. In this study, we take advantage of generic computational techniques to enable the dynamical behaviour of complex cellular network models.
A first approach, called "model verification", enables the formalisation and the automated verification of validation criteria for whole models or selected subparts, thereby greatly facilitating model development and correction.
A second approach, "value percolation", enables the computation of the impact of specific environmental or genetic conditions on model dynamics.
We apply these methods to the analysis of the pathways involved in checkpoint inhibitor blockade, a domain of cancer immunotherapy under active scrutiny.
The proposed methods and models will soon be made available in the all-inclusive CoLoMoTo Docker image, which provides a reproducible modeling environment, and in an interactive companion notebook.

Tom Freeman: A graphical and computational model of the renal mammalian circadian clock

The circadian clock schedules an organism's internal physiology and behaviour to function at the appropriate time of day. In mammals, the core molecular components of the clock and the interactions between them, are highly conserved between all tissues and cells, but their phases and the downstream effects are generally tissue- or cell-specific. In the kidney, the circadian clock plays a pivotal role in regulating daily fluctuations in blood pressure primarily through the modulation of sodium transport and extra-cellular fluid volume. Perturbations of this rhythm, particularly the nocturnal dip, confer increased risk for cardiovascular and renal disease. To better appreciate the circadian biology of the kidney, we have sought to analyse the diurnal pattern of gene expression in this organ and relate these observations to the core components of the circadian clock. First, we examined transcriptomics data (CircaDB) describing the variation in gene expression in the murine kidney over a 48-hour period, identifying those genes that exhibited a diurnal pattern of expression. These data we used as a reference to the activity of the core clock components within the kidney and the transcriptional networks they regulate. To model the circadian clock pathway, existing models of the mammalian clock were examined and the primary literature was searched for studies describing its molecular components and the interactions between them. Using this information, a comprehensive graphical model of the pathway constructed using the modified Edinburgh Pathway Notation scheme1. This was assembled in such a way to allow its parameterisation for simulation experiments using a stochastic Petri net-based approach2. mRNA levels CircaDB were used as a proxy for protein levels to define initial conditions and 'delay motifs' introduced to modulate activity flow over time. Following empirical testing the model was further parameterised such that the simulated 'expression' of core components closely matched their observed activity in the mouse kidney. The result of this work is a detailed graphical and computational model, which summarises the current literature of the molecular interactions that regulate the mammalian circadian clock. It encapsulates the interactions between 69 molecular species and together with other motifs, the model contains 2013 nodes and 2100 edges. Following parameterisation, simulations using the model recapitulated the transcriptional activities of the regulated components in the mouse kidney. Virtual knock-out experiments performed on the model were shown to reflect experimental data. It also identified points at which canonical clock genes may integrate with downstream genes likely to affect blood pressure and other aspects of kidney function. We believe that the model provides new insights into the complexity and function of this most central of physiological pathways.

Paul Thomas: Gene Ontology Causal Activity Modeling

Gene Ontology (GO) annotations are the most comprehensive structured representation of gene function, and are widely used in the interpretation of genome-wide experimental data. However, because an individual GO annotation associates a single gene product with a single GO term, it is only a partial description of gene function, which limits the expressiveness of annotations and their application in computational analysis of experimental data. To address this limitation, we have developed a framework, GO Causal Activity Modeling (GO-CAM), for linking multiple GO annotations into an integrated model

of a biological system. GO-CAM provides a structured framework for GO annotation, similarly to how the GO ontology has long provided a framework for a controlled vocabulary of gene function. GO-CAM supports modeling at multiple levels, from individual gene products to complex regulatory and metabolic pathways, and can be applied in network analysis and systems biology modeling, or converted into standard GO annotations for traditional GO-based analyses.
I will discuss the GO-CAM formalism, and show the Noctua Modeling Tool that is used by GO Consortium curators to create GO-CAM models from existing GO annotations, as well as from scratch.

Anna Niarakis: Automated inference of annotated Boolean models from molecular interaction maps using CaSQ

Introduction: Biological processes rely on the concerted interactions and regulations of thousands of molecules capable of forming complex networks. Disruption and dysregulation of these networks can lead to disease. Molecular interaction maps have emerged as a useful way of representing biological mechanisms, based on information mining and human curation (Ostaszewsky et al. 2018). They can be analysed topologically or serve as templates for visualising omics datasets. Their static nature does not allow for studying the emerging behaviour of the system under different conditions. Dynamical modelling should be employed for in silico simulations. We develop CaSQ (Singh et al, 2019a, under preparation), a tool that infers preliminary Boolean rules based on topology and semantics of molecular interaction maps. We apply this method to build a large scale dynamical model for Rheumatoid Arthritis (RA), a complex disease of unknown aetiology.
Methods: We have used a state-of-the-art molecular interaction map for Rheumatoid Arthritis (Singh et al, 2018, Singh et al, 2019b, under preparation). Based on the topology, semantics and annotations of the RA map, we have defined simplification rules and logical formulas that we have compiled in a tool, CaSQ. We have used CaSQ to produce executable files of the RA map to study RA fibroblast (RASF) activation. Additionally, we have tested CaSQ with a number of existing molecular maps that differ in size, complexity and the use of standards to evaluate the applicability of our method.
Results: Direct conversion of the RA map results in a disease- but not cell-type specific Boolean model. Our approach is to extract RASF specific subnetworks, use CaSQ to infer the Boolean model and analyse it using the modelling platform Cell Collective (Helikar et al, 2012). CaSQ can handle various maps and produce models in a qualitative Systems Biology Markup Language (SBML-qual) format while references, annotations and layout are retained facilitating interoperability and model reusability. Comparison of CaSQ-inferred models against manually derived ones from corresponding maps show that CaSQ is capable of capturing the dynamics of complex networks.
Conclusions: We have successfully converted the RA map and others to executable Boolean models, using CaSQ, a tool that bridges the gap between static and dynamic representations, offering a major drive in the modelling of complex disease networks.

Session 3 – Tools (I)

Tomas Helikar: Cell Collective - Enabling accessible and collaborative construction and analysis of comprehensive and annotated models

Cell Collective is a freely available, web-based, interactive computational modeling platform for the collaborative construction, simulation, and analyses of large-scale dynamic (logical) models of biological and biochemical processes. Cell Collective also contains nearly 100 public, peer-reviewed computational network models of various biological and biochemical processes, such as gene regulatory networks, signal transduction, and cell cycle in organisms ranging from bacteria and viruses to yeast, flies, plants, to humans. Models in Cell Collective can be created either de novo or they can be imported using the SBML-qual standard.

Importantly, not all users (existing and potential) and contributors to computational models are trained in mathematics, computer programming, and/or computational modeling. As such, a major design focus of the Cell Collective platform has been to enable computational modeling to play an integral role in experimental hypothesis generation and testing by making the platform accessible to users that may or may not have a computational background. To enable the scientific community to efficiently use existing models and stimulate model re-use and expansion, every component and interaction is annotated to track the biological data used from scientific literature to build the model. Cell Collective supports real-time simulations, enabling users to assess the dynamics of models under various what-if scenarios (e.g., perturbations, drug effects, etc.). Cell Collective offers a suite of built-in tools for model analysis, including dose-response analysis, environment sensitivity, topology, and feedback loops.

Cell Collective provides an environment that enables collaborative construction, annotation, and simulation/analysis of computational models directly in the platform. This means that a broad community of researchers who are experts in different aspects of the model areas (biology) can collaborate to construct and validate models that are meaningfully comprehensive, detailed and accurate. Specifically, each model in Cell Collective can be shared with specific collaborators or published for the broader scientific community. Models are accessible in Cell Collective or can be downloaded in SBML format for analyses in other software tools.

Finally, With data being generated by scientists at a staggering rate in the course of studying biological systems, computational modeling and simulations have emerged as integral to undergraduate life sciences education. Life sciences students now need skills to reason conceptually, mechanistically, and quantitatively to answer emerging life science questions. As such, in designing Cell Collective to be broadly accessible, the technology has also become a full-scale educational tool for life sciences courses. Life sciences students, too, can now learn about biological processes by creating, simulating, and interpreting computational models of complex systems. The learning approach delivered through Cell Collective has been used by thousands of students in institutions and courses worldwide: high school, introductory college, specialized upper- undergraduate, and graduate levels, including in undergraduate immunology and microbiology courses, cancer biology courses and introductory biology courses.

Laurence Calzone + Gaultier Stoll: MaBoSS ecosystem

MaBoSS (Markovian Boolean Stochastic Simulator) is a tool for simulating logical models with continuous time Markov processes. The outputs of MaBoSS stochastic simulations provide probabilities for each state of the model, reflecting the diversity of a cell population and informing on the proportion of the diverse cellular states for a variety of cell conditions. Over the years, some biological features were added to the initial purpose of MaBoSS. These features include death, division, intercellular interactions, intracellular impact of the microenvironment, or the effect of perturbations. For this purpose, several tools were developed using MaBoSS as the core simulator. Among them, Ensemble modeling, UPMaBoSS and PhysiBoSS are briefly introduced here.
Ensemble modeling is a methodology that simulates and studies sets of logical models that share the same variables, have common biological constraints, but different logical rules. It explores the possible model dynamics and highlights properties of the networks' structure. UPMaBoSS studies the dynamical behavior of a cell population by considering cell death, cell division and intercellular communication. In practice, a MaBoSS simulation is halted at regular intervals and the status of each cell of the population is updated. The tool allows to study the dynamics of a population in response to different perturbations (mutations, drug treatments, etc.). PhysiBoSS is a tool based on an agent-based formalism. Each agent is a logical model run with MaBoSS. In this case, physical and environmental features are considered in the cellular response to external cues. PhysiBoSS produces a 2-D or a 3-D representation of a cell population subjected to different cellular conditions.
We will present the different ways to run these tools with a running case study of a model of cell fate decision: through command line; through the CoLoMoTo Interactive Notebook by using the python library pymaboss; and through the user-friendly web interface WebMaBoSS.

Vasundra Touré: The Minimum Information about a Molecular Interaction Causal Statement (MI2CAST): a guideline for the management of molecular causal interaction

In Systems Biology, regulatory process networks are built to reflect how components in cell fate decision systems are interconnected and behave. A considerable amount of knowledge provided by different public resources is available in the form of large biological networks depicting e.g. metabolic reactions, signaling cascades and gene regulatory events. We aim at using this information by disassembling those networks into their most basic regulatory network motifs, called "causal statements". A causal statement is describing a directed interaction where a source entity (regulator) has an influence over the quantity or the activity of a target entity (regulatee). By looking at the core interactions occurring among entities, the understanding of the mechanisms they enable in biological regulations could be improved, and conversely, by specifying the elements of regulatory reactions in sufficient detail new signaling networks accommoding alternative cellular behaviour may be composed easily from its network constituents. Once causality is observed, the next challenge is to represent and archive it so that it can be shared, reused and reconstituted in causal networks by computers and humans alike. At present, various representations of causal relationships between biological components are used in a variety of resources.

However, they capture different aspects of contextual details about causal interactions. We propose the minimum information about a molecular interaction causal statement (MI2CAST) to formalize the information that ideally should be captured when representing causal interactions through an unambiguous data description. This reporting guideline should be considered as a checklist that can be followed in curation processes and to consult when building curation templates that accommodate capturing the essential contextual information about a causal relationship. The aim is to ensure clarity, uniformity and reusability of the data across resources.

Session 4 – Tools (II)

Julio Saez Rodriguez: Integrating knowledge and experimental data to build context-specific logic models

Dynamic logic models are a powerful tool to understand biological networks.Over the years, we have developed methods and tools to apply this logic formalism to build context-specific models, with a focus on signalling networks and the use of data obtained upon perturbation. Our general pipeline involves obtaining existing prior knowledge on pathways from available public resources using our tool OmniPath (www.omnipathdb.org), building a logic model from this prior knowledge, and training it to data with our tools CellNOpt (for targeted readouts - www.cellnopt.org), Phonemes (for untargeted mass spectrometry proteomics), and CARNIVAL (for gene expression). Formalism variants allow us to handle variables as either Boolean (binary) or continuous and even to be casted as differential equations. I will describe recent methodological developments, including extensions to model metabolic regulation. I will illustrate their utility in cases of biomedical relevance, in particular to improve our understanding of cancer and develop novel therapeutic opportunities. I will also present how we use Omnipath to capture information from annotations from 36 different resources.

Aurélien Naldi: The CoLoMoTo Interactive Notebook: Accessible and Reproducible Computational Analyses for Qualitative Biological Networks

Analysing models of biological networks typically relies on workflows in which different software tools with sensitive parameters are chained together, many times with additional manual steps.
The accessibility and reproducibility of such workflows is challenging, as publications often overlook analysis details, and because some of these tools may be difficult to install, and/or have a steep learning curve.
The CoLoMoTo Interactive Notebook provides a unified environment to edit, execute, share, and reproduce analyses of qualitative (logical) models of biological networks. This framework combines the power of different technologies to ensure repeatability and to reduce their learning curve. The CoLoMoTo Interactive Notebook currently provides access

to software tools including GINsim, BioLQM, Pint, MaBoSS, and Cell Collective for the modelling and analysis of Boolean and multi-valued networks. More tools will be included in the future. We developed a Python interface for each of these tools to offer a unified and seamless integration and to ease the chaining of complementary analyses. Computational workflows can be edited through a web interface based on the Jupyter notebook, enabling the inclusion of textual annotations, along with the explicit code to execute, as well as the visualisation of the results. The resulting notebook files can then be shared and re-executed in the same environment.
The framework is distributed as a Docker image with the tools ready to use without any installation step besides Docker, which can run on Linux, macOS, and Microsoft Windows systems. The docker image further provides several tutorial notebooks and use cases, with applications related to immunology and cell-fate decisions. Further information and installation instructions are available on http://colomoto.org/notebook/.

Eugenia Oshurko: KAMIStudio

Modelling cellular signalling remains a challenging task due to the complexity of the underlying dynamical systems emerging from an enormous number of interactions performed by complex heterogeneous agents. It is not only computationally hard to simulate and analyse such systems, even writing down their models is in itself a highly non-trivial problem due to the combinatorial explosion in the number of agent species and the fragmentary nature of signalling knowledge. Rule-based modelling languages (such as Kappa and BioNetGet) made some progress in tackling these challenges. Namely, they solve the combinatorial complexity problems inherent to signalling models, allow incremental model building and even propose some tools for static and causal analysis of dynamical models.
However, they remain unsuitable for collation, maintenance and reuse of signalling knowledge. While different rules express conditions for individual interactions between agents, these interactions are not necessarily independent and may share interaction mechanisms. This poses what we call the update problem, i.e. an update of knowledge on an interaction mechanism may require manual identication and respective update of all the instances of this mechanism (rules expressing them). Moreover, the representation proposed by the rule-based modelling languages does not
capture interaction conditions based on the presence or absence of some conserved protein domains or specific key residues, which is essential for understanding and inferring the interaction capabilities of different protein variants (e.g. splice variants, mutants). The above-mentioned issues compromise the reuse of the knowledge expressed with rule-based models and its adaptation to different contexts.
In this talk we will present the KAMIStudio environment based on the KAMI bio-curation framework. KAMI aims to decouple knowledge curation from model building by de-contextualizing protein-protein interaction (PPI) knowledge and articulating the notion of an interaction mechanism. We will briefly present the two types of knowledge bodies used in KAMI: corpora containing de-contextualized knowledge and models containing knowledge instantiated in given contexts, as well as the graph-based knowledge representation system used to accommodate corpora and models. We will present the main features of the KAMIStudio environment that can be used for semi-automatic curation of

large corpora of cellular signalling knowledge including: interactive visualization of knowledge stored in corpora and models; input of individual PPIs to a corpus resulting in the automatic aggregation of the new knowledge to the corpus; an interface for specifying protein variants (isoforms); automatic instantiation of corpora into signalling models using protein variants; and automatic generation of Kappa scripts from models that can be further used to study the dynamics of the modelled systems.

Workshop Schedule

Time	Details
9h00-9h10	**Welcome and introduction to the workshop**
9h10-10h40	**Session 1 – Data/model curation/annotation and available repositories –** Chair Denis Thieffry – Anna Niarakis
9h10-9h30	**Martin Kuiper:** Towards a curation platform for causal interaction statements.
9h30-9h50	**Marek Ostaszewski:** BioKB and MINERVA: a workflow for curation and quick prototyping of annotated knowledge repositories
9h50-10h10	**Sheriff Malik:** Curation and annotation of models in BioModels repository promotes reproducibility and reusability
10h10-10h30	**Cristina Casals:** SysVasc Prior Knowledge Network: An example of biocuration for Boolean modelling
10h30-11h00	Coffee/tea break
11h00-12h30	**Session 2 – Community standards development and interoperability/reusability –** Chair Aurélien Naldi – Laurence Calzone
11h00-11h20	**Denis Thieffry:** Computational verification of large logical models - application to the prediction of T cell response to checkpoint inhibitors
11h20-11h40	**Tom Freeman:** A graphical and computational model of the renal mammalian circadian clock
11h40-12h00	**Paul Thomas:** Gene Ontology Causal Activity Modeling
12h00-12h20	**Anna Niarakis:** Automated inference of annotated Boolean models from molecular interaction maps using CaSQ

12h20-12h30	Wrap up of morning sessions
12h30-13h30	Lunch break with coffee fix
13h30-14h00	Round table – Discussion - Exchanges
14h00-15h00	**Session 3 – Tools (I) –** Chair Julio Saez Rodriguez
14h00-14h20	**Tomas Helikar:** Cell Collective
14h20-14h40	**Laurence Calzone + Gaultier Stoll:** MaBoSS ecosystem
14h40-15h00	**Vasundra Touré:** The Minimum Information about a Molecular Interaction Causal Statement (MI2CAST): a guideline for the management of molecular causal interaction
15h00-16h00	**Session 4 – Tools (II) –** Chair Tomas Helikar
15h00-15h20	**Julio Saez Rodriguez** CellNOpt
15h20-15h40	**Aurélien Naldi:** The CoLoMoTo Interactive Notebook: Accessible and Reproducible Computational Analyses for Qualitative Biological Networks
15h40-16h00	**Eugenia Oshurko:** KAMIStudio
16h00-16h15	Discussion – Concluding remarks